Method for Accurate Unsupervised Cell Nucleus Segmentation

Pascal Bamford[1] and Brian Lovell[1,2]
[1]Cooperative Research Centre for Sensor Signal and Information Processing
[2]School of Computer Science and Electrical Engineering
University of Queensland, Australia 4072
P.Bamford@cssip.uq.edu.au, lovell@csee.uq.edu.au

Abstract - **To achieve the extreme accuracy rates demanded by applications in unsupervised automated cytology, it is frequently necessary to supplement the primary segmentation algorithm with a segmentation quality control system. The more robust the segmentation strategy, the less severe the data pruning need be at the segmentation validation stage. These issues are addressed as we describe our cell nucleus segmentation strategy which is able to achieve 100% accurate segmentation from a data set of 19946 cell nucleus images by automatically discarding the most difficult cell images. The automatic quality checking is applied to enhance the performance of a robust energy minimisation based segmentation scheme which already achieved a 99.47% accurate segmentation rate.**

Keywords — **cell, cytology, image, segmentation, robust**

I. Introduction

Machine vision systems for the unsupervised automation of otherwise manual tasks usually require image processing components with exceptionally high accuracy rates. This is especially true in the biomedical domain where failures result in mis-diagnoses. The fact that research is still continuing on the development of a cervical cancer screening machine despite projects being initiated in the 1950's is perhaps a good indication of the magnitude of the development effort required to go from an algorithm obtaining "good" results on a small test data set, to one obtaining acceptable levels of accuracy in a real environment. The main difficulty with this application has been identified as the robust segmentation of cells and cell nuclei. Indeed, Bengtsson [5] says that segmentation stage is "the key to a working machine" echoing the sentiments of Gonzalez and Woods [9] that, "effective segmentation rarely fails to lead to a successful solution." Many algorithms have been proposed in the past with varying degrees of success, but just as important as a high accuracy rate is knowing when a failure has occurred as '... an erroneously segmented cell is much worse than a rejected cell' [13].

Many researchers have included artefact and incorrect segmentation rejection schemes in their algorithms. MacAulay used a post-processing step after segmentation to remove potential artefacts based on shape and appearance that was capable of detecting some of the incorrectly segmented nuclei [11]. Nordin describes an algorithm that is able to report failures at various stages of the segmentation process, as well as a separate artefact rejection algorithm [13]. McKenna made use of a neural network to select potential nuclei in scenes for subsequent segmentation. He noted that a post-processing stage would also be necessary to filter out "erroneously detected objects" [12].

A common theme in these techniques is the need for a separate quality control process to view the output of the segmentation and typically these apply shape and appearance criteria to classify the results as either "pass" (looks like a cell) or "fail" (doesn't look like a cell). We have developed a segmentation strategy that not only employs a segmentation algorithm with much higher performance than previously reported [4], but which also provides a confidence measure in the resulting segmentation without explicit reference to shape and appearance criteria for quality check purposes.

II. The Segmentation Stage

For a full explanation of the underlying segmentation technique, the reader is referred to [4] and [1]. The method is based on energy minimisation techniques and is summarised here only to introduce the development of the subsequent quality checking strategy.

A. Energy Minimisation Implementation

The use of active contours in bio-medical applications is well established, but it is well known that these methods tend to suffer from local minima, initialisation, and stopping criteria problems. Fortunately global minimum energy

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle Method for Accurate Unsupervised Cell Nucleus Segmentation	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Cooperative Research Centre for Sensor Signal and Information Processing University of Queensland, Australia 4072	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
5

searching methods have been found to be particularly effective in avoiding local minima problems due to the presence of the many artefacts often associated with medical images [6][7][8]. Here, a dynamically programmed search method was implemented based upon a suggestion in [10]. A search space is first set up within the image, bounded by two concentric circles centralised upon the approximate centre of the nucleus found by an initial rough segmentation technique (*e.g.*, converging squares algorithm). This search space is sampled to form a circular trellis by discretising both the circles and a grid of evenly-spaced radial lines joining them (figure 1).

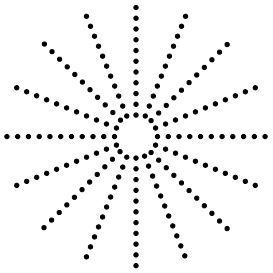


Figure 1. Discrete search space

Every possible contour that lies upon the nodes of the search space is then evaluated and an associated energy or cost function is calculated. This cost is a function of both the contour's smoothness and how closely it follows image edges. The relative weighting of the cost components is controlled by a single regularisation parameter, $\lambda \in [0, 1]$. By choosing a high value of λ, the smoothness term dominates, which may lead to contours that tend to ignore important image edges. On the other hand, low values of λ allow contours to develop sharp corners as they attempt to follow all high gradient edges, even those which may not necessarily be on the desired objects edge. Once every contour has been evaluated, the single contour with least cost is chosen as the global solution. The well-known Viterbi algorithm provides an efficient method to find this global solution as described in [4].

B. Segmentation Performance

A data set of 19946 Pap stained cervical cell images was available for testing. These images were of the order of 128x128 pixels, quantised to 256 gray levels and each contained a single nucleus.

The single parameter λ that controls the behaviour of the algorithm, was empirically chosen to be 0.7 after trial runs on a small sub-set of the images. This sub-set was made up of 141 known 'difficult' images from previous studies [4][3], augmented by a random sample of 269 images from the remaining data set. This careful data selection was necessary as previous experience showed that for the majority of images, the resulting segmentation was fairly insensitive to the choice of λ, making the choice of optimum value difficult. Nevertheless, more demanding images require some adjustment to the parameter to achieve correct segmentation. The effect of the choice of λ on segmentation accuracy on this trial set is shown by the graph of figure 2.

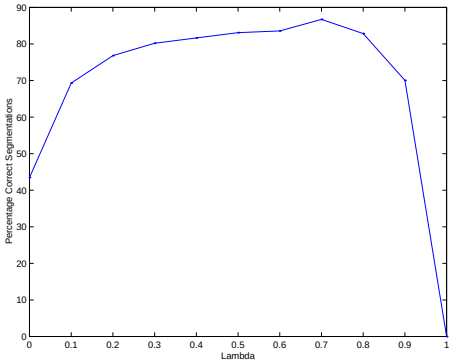
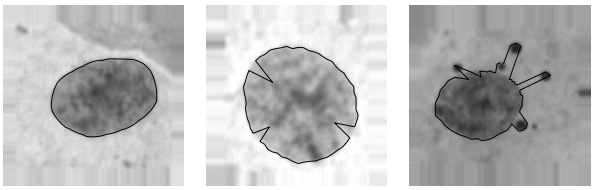


Figure 2. Plot of percentage of correct segmentations against λ for a set of images consisting of known 'difficult' images and randomly selected images.

With λ set at 0.0, the smoothness constraint is completely ignored and the point of greatest gradient is chosen along each search space radius. Previous studies [3] have shown that for approximately 65% of images, all points of greatest gradient actually lie upon the nucleus cytoplasm border (figure 3(a)), so these cell images will be correctly segmented.

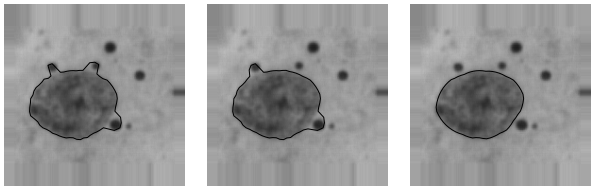


(a) (b) (c)

Figure 3. $\lambda = 0.0$. a) Largest gradients occur on the nucleus border, b) darkly stained chromatin generates largest gradients, c) dark artefacts generate largest gradients.

For the remaining 35% of images, a large gradient due to an artefact or darkly stained chromatin will draw the contour away from the desired border (figures 3(b)&(c)). As λ increases, the large curvatures present in these configurations become less probable (figure 4).

The graph shows a value of $\lambda = 0.7$ as the most suitable for these particular images. Every image in the data set was



(a) (b) (c)

Figure 4. The effect of increasing λ. a) $\lambda = 0.1$, b) $\lambda = 0.2$, c) $\lambda = 0.5$.

then segmented at $\lambda = 0.7$ and the results verified by eye. Of the 19946 images, 99.47% were found to be correctly segmented. Three main classes of failure were identified. Eighty seven of the failures were due to the nuclei lying close to the cytoplasm boundary. As the background cytoplasm boundary contrast is much greater than that of the nucleus cytoplasm boundary, the contour tended to lie upon the former very low image energy area (high gradient edges). Fourteen of the failures were caused by the inappropriate choice of λ for that individual image (they all subsequently produced correct segmentations with different values of λ.) The remaining four images were found to fail at all attempts. The failures due to the presence of the background in the nucleus images are preventable through careful design of a prior cell-finding stage [2]. Here, the cytoplasm background boundary is known and can therefore be prevented from appearing in the nucleus images. The detection of the other classes of failure is therefore the major issue.

III. Development of an Error Checking Framework

Despite the exceptionally high accuracy rate that the global minimum searching contour method achieves, there is still a possibility of sample contamination from the few failures that do occur. In order to prevent this, the need would still exist for a human to view the output of this stage, undermining its utility in a practical system. The remainder of this paper therefore concerns itself with the development of a framework that further increases the accuracy of a potential system.

A. Lambda Sensitivity

For the majority of relatively simple images with little ambiguity in the true location of the nuclear boundary, the final segmentation can be fairly insensitive to λ over a wide range of values (figure 5).

By contrast, 'difficult' images (even for humans) produce very different contours depending upon the choice of λ (figures 4 and 6).

These images usually contain artefacts near or on the nuclear boundaries that make the 'true' border hard to find. These examples show that no single value of λ is capable of accurately segmenting all of the images. Therefore, rather than segment the images at one value of λ and use a post-process to reject possible failures, we are interested in viewing the output of the algorithm for various values of λ in order to detect stability as a measure of confidence in the resulting segmentation.

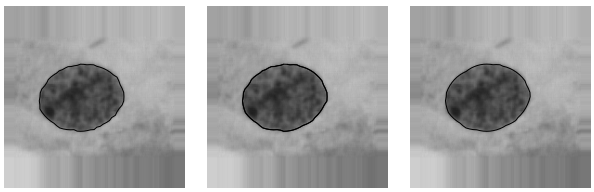
B. Error Checking

The graph of figure 2 shows monotonically increasing segmentation accuracy for $0.0 < \lambda < 0.7$. In fact, from the data it was observed that the set of correct segmentations at λ_1 was a strict subset of the set of correct segmentations at λ_2 where $\lambda_1 < \lambda_2 < 0.7$. Therefore, by segmenting an image at the high probability of correct segmentation value of $\lambda = 0.7$ and then again at $\lambda = 0.0$, similarity between the two contours indicates a high level of contour stability (figure 5). This image is then classified as a 'very easy' image to segment and for convenience labelled "level 0". Lack of similarity leads to a comparison of the contour at $\lambda = 0.7$ with the contour at $\lambda = 0.1$. Similarity leads to a classification of level 1 and so on.

This classification method suggests a means to discard incorrect segmentations. For example, if we keep only level 0 (very easy) cell images, we discard approximately a third of the data set, but achieve a 100% correct segmentation rate on those retained [3].

IV. Fine Tuning

In order to pursue this method, the data set was split into two sets: $\mathcal{F}$, Those images that been incorrectly segmented at $\lambda = 0.7$ (105 images) and $\mathcal{C}$, those that had been correctly segmented (19841 images). Statistics were then measured for each level by comparing the segmentation result at $\lambda =$



(a) (b) (c)

Figure 5. Example of an image that is stable over a range of λ. (a) $\lambda = 0.1$, (b) $\lambda = 0.5$, (c) $\lambda = 0.7$.

Level	0	1	2	3	4	5	6
Threshold	4.85	3.20	2.45	0.79	0.79	0.79	0.79

Table 1. Minimum MAD thresholds for the detection of every element in $\mathcal{F}$ (incorrect segmentation at $\lambda = 0.7$ on the test data set) for levels 0 - 6.

0.7 with those at $\lambda = 0.0, 0.1, ..., 0.6$ for every image in both sets.

As the contours to be compared were the result of the same algorithm and indeed the same search space in the image, the comparison between any two contours is trivial. The distance between each chosen point on each of the search space radii (figure 1) for each contour was calculated and the maximum absolute deviation (MAD) evaluated.

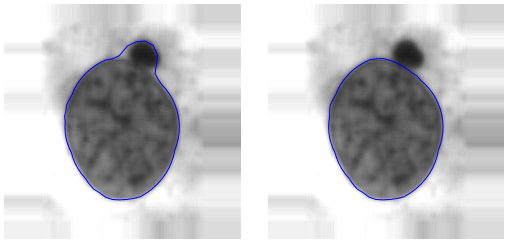
A cumulative plot of the percentage of the set $\mathcal{F}$ against MAD for level zero (comparison between contours at $\lambda = 0.7$ and $\lambda = 0.0$) is shown in figure 7.

This graph shows that for level 0, a MAD threshold of 4.84 pixels would detect every failed segmentation. In a similar manner, it is possible to establish thresholds for each level so that the detection of every failed segmentation in this database is guaranteed (table 1).

The thresholds decrease with increasing level. This is expected as closer values of λ are compared at higher levels. The values then taper to a limit of 0.79 pixels as this is the distance between two adjacent radial points on the discrete search space.

In order to establish the effect of setting such thresholds on $\mathcal{C}$, the percentage of $\mathcal{C}$ that would be discarded against MAD threshold for level zero is shown in figure 8.

Therefore, by setting a threshold of 4.84 pixels and rejecting any segmentation with a greater MAD, 40.2% of the correct segmentations would be discarded. This procedure may be repeated for each level, using the thresholds previously calculated. The percentage of C that falls above the threshold for each level (i.e. a good segmentation being discarded) against MAD is shown in figure 9.



(a) (b)

Figure 6. Example of an image that is not stable over a range of λ. (a) $\lambda = 0.5$, (b) $\lambda = 0.7$

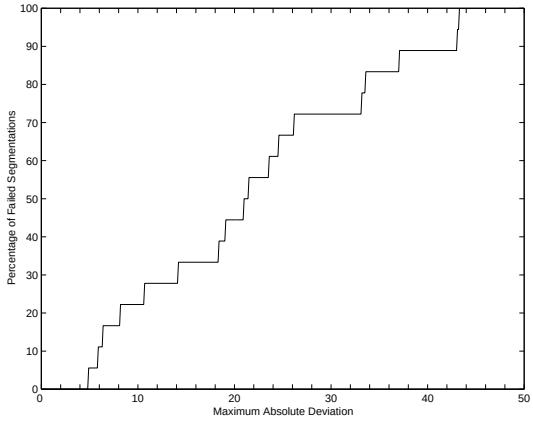


Figure 7. A plot of the percentage of the elements of $\mathcal{F}$ (incorrect segmentation at $\lambda = 0.7$ on the test data set) against measured MAD for level 0.

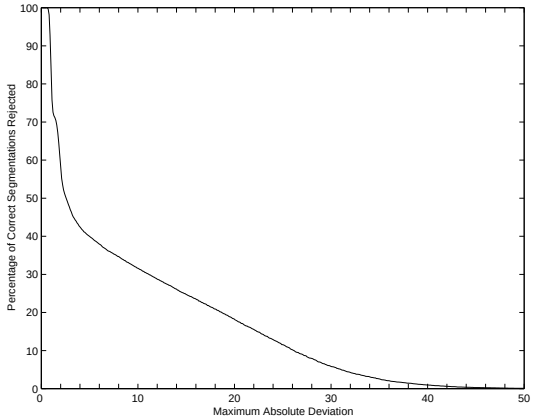


Figure 8. A plot of the percentage of elements of $\mathcal{C}$ (correct segmentations at $\lambda = 0.7$ on the test data set) rejected against MAD threshold for level 0.

Although a harsher threshold is used at level 1 than at level 0, fewer correct segmentations are discarded. This is due to the absence of any smoothness constraint at $\lambda = 0.0$ which leads to the wild deviations such as those shown in figure 3. However, the small smoothness contribution at $\lambda = 0.1$ corrects many of these deviations resulting in the large drop in average MAD (table 2).

Therefore by running at level 2, it is possible to detect every failure and only discard 10.78% of the correct segmentations.

V. Conclusions

By analysing the modes of failure of a highly successful cell nucleus segmentation algorithm, an error checking

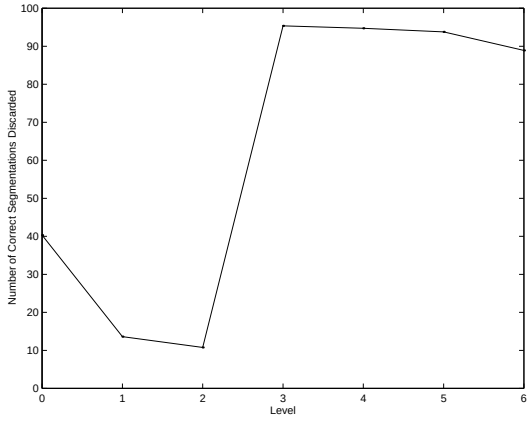


Figure 9. Plot of the percentage of elements of $\mathcal{C}$ (correct segmentations at $\lambda = 0.7$ on the test data set) rejected at each level using the thresholds of table 1.

Level	0	1	2	3	4	5	6
Average MAD	8.90	2.78	1.91	1.56	1.35	1.22	1.07

Table 2. Average MAD for levels 0 - 6.

framework was implemented that was capable of detecting every failure. The algorithm parameter $\lambda$was first empirically tuned for the data set to obtain best segmentation accuracy. It was then observed that different values of λ obtained different solutions for difficult images, but simple images generated stable solutions. Therefore, by varying λ this stability could be detected. A decision to reject or accept the segmentation was then made, based upon measured thresholds for each level. For the data set of 19946 images, it was found that by comparing the resulting contours at values of $\lambda = 0.7$ and $\lambda = 0.2$, and rejecting the segmentation if the maximum absolute deviation (MAD) between the contours was greater than 2.45 pixels, every failure could be detected whilst only discarding 10.78% of the correct segmentations. In this study, only values of λ with a resolution of 0.1 have been considered. It is possible that by increasing this resolution in the region of interest (i.e. near 'level 2' operation) and repeating the exercise, a further increase in the performance could be achieved. Naturally, the parameters and results that have been reported are optimised not only for one type of image but also for the hardware configuration that was used to capture them. Current work involves the incorporation of the proposed system into a fully automated *Cytometer* (an automatic imaging system) using the same methodology to achieve optimal performance for that hardware. This allows for much more extensive analysis of the proposed methods through the accessability of a greater amount of data. This result has great potential for the implementation of unsupervised cancer screening devices using methods where only a representative sample of cells is required. Preliminary studies have shown strong possibilities for the non-invasive early detection of lung and other forms of cancer.

Finally, The 'rejected' cells have simply been labelled as such. These could be interpreted as having been 'flagged' by the algorithm as problematic, requiring processing by a higher level (*e.g.*, to invoke a different algorithm etc.) By achieving such high accuracy rates and confidence in the segmentation stage, the following feature extraction and classification processes can only become more robust.

REFERENCES

[1] P. Bamford. *The Segmentation of Cell Images with Application to Cervical Cancer Screening*. PhD Thesis, The University of Queensland, 2000.

[2] P. Bamford and B. Lovell. A water immersion algorithm for cytological image segmentation. In *Proceedings of the APRS Image Segmentation Workshop*, pages 75–79, University of Technology Sydney, Sydney, December 1996.

[3] P. Bamford and B. Lovell. Improving the robustness of cell nucleus segmentation. In P. H. Lewis and M. S. Nixon, editors, *Proceedings of the Ninth British Machine Vision Conference, BMVC '98*, pages 518–524, University of Southampton, England, UK, September 1998.

[4] P. Bamford and B. Lovell. Unsupervised cell nucleus segmentation with active contours. *Signal Processing Special Issue: Deformable Models and Techniques for Image and Signal Processing*, 71(2):203–213, December 1998.

[5] E. Bengtsson. The measuring of cell features. *Analytical and Quantitative Cytology*, 9(3):212–217, June 1987.

[6] L. D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1131–1147, 1993.

[7] C. A. Davatzikos and J. L. Prince. An active contour model for mapping the cortex. *IEEE Transactions on Medical Imaging*, 14(1):65–80, 1995.

[8] D. Geiger, A. Gupta, L. Costa, and J. Vlontzos. Dynamic programming for detecting, tracking, and matching deformable contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):294–302, 1995.

[9] R. C. Gonzalez and R. E. Woods. Digital Image Processing Addison-Wesley, Reading, Massachusetts, 1993.

[10] S. R. Gunn. *Dual Active Contour Models for Image Feature Extraction*. PhD thesis, University of Southampton, May 1996.

[11] C. E. MacAulay. *Development, Implementation and Evaluation of Segmentation Algorithms for the Automatic Classification of Cervical Cells*. PhD thesis, University of British Columbia, August 1989.

[12] S. J. McKenna. *Automated Analysis of Papanicolaou Smears*. PhD thesis, University of Dundee, October 1994.

[13] B. Nordin. *The Development of an Automatic Prescreener for the Early Detection of Cervical Cancer: Algorithms and Implementation*. PhD thesis, Uppsala University, 1989.